Visualization at the NCCS

NCCS USERS MEETING

Sean Ahern, ORNL

Mar 28, 2007

Outline

- **Visualization efforts and team**
- **Hardware resources**
  - Visualization cluster
  - EVEREST display
- **Visualization tools**
- **Data access issues and solutions**

Many Coordinated Efforts

The visualization effort within the NCCS strives to deliver visual data analysis tools and capabilities to our customers. It includes many sub-efforts to successful assist their scientific mission:

- Support visualization tools
- Convert data
- Perform statistical analyses
- Produce publication images
- Produce movies
- Highlight science successes to visitors
- Explore new data exploration techniques
- Write custom modules
- Write custom visualization tools
- Parallel analysis support
- Large display support

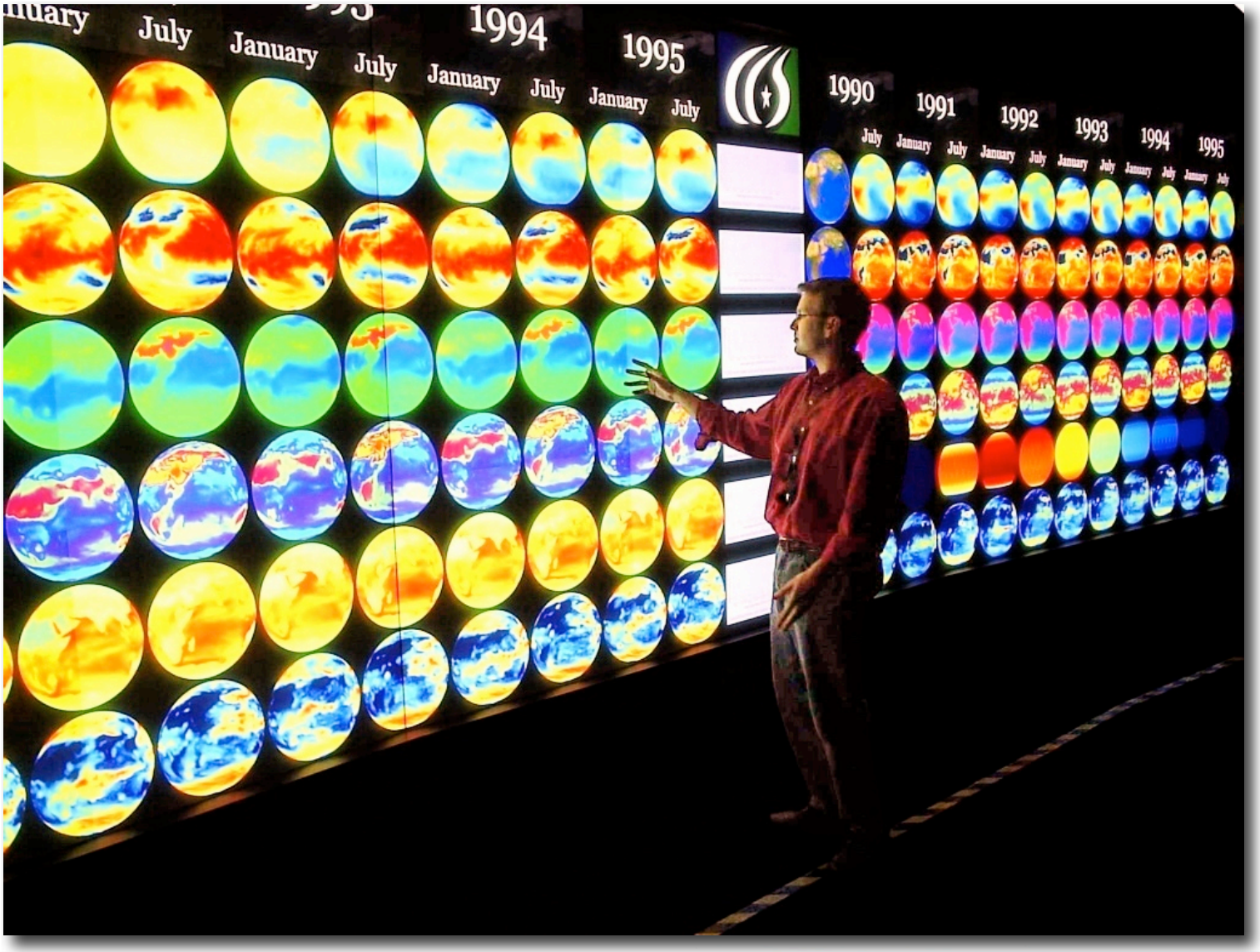
Visualization Team

Sean Ahern
Jamison Daniel
Ross Toedte
George Ostrouchov
Jeremy Meredith

Hawk Visualization Cluster

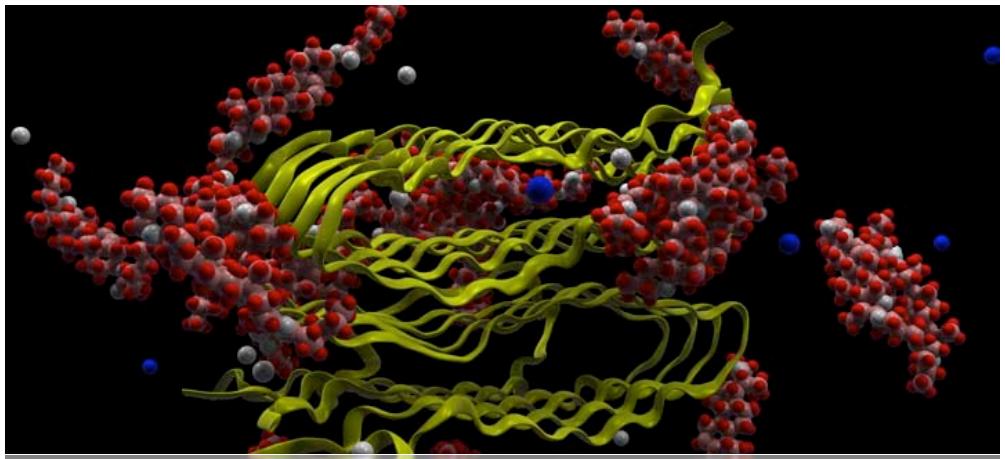
- Dedicated visualization cluster
- 58 nodes
- Dual opterons, 1.6 MHz
- 116 gigs of memory
- Quadrics Elan3 and GigE interconnect
- NVIDIA 5900 and NVIDIA QuadroFX 3000G GPUs
- High-speed connection to 10 GigE infrastructure
- Pending Lustre integration
- 14 nodes “dedicated” to EVEREST
- Batch system managed by SLURM
- Firewall exception for hawk.ccs.ornl.gov, providing easy remote access to parallel vis resource

EVEREST


July
January
July
1994
January
July
1995
January
July
1990
1991
1992
1993
1994
1995

Science Communication

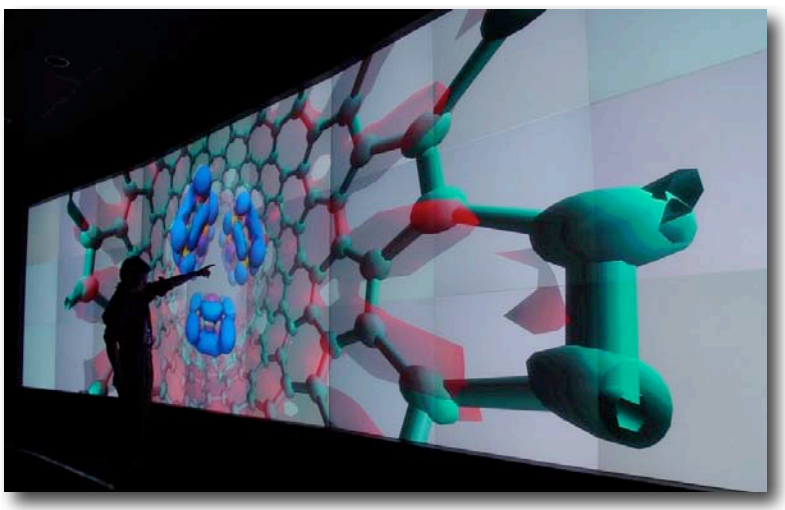
- **We give 400-500 tours a year to external visitors**
- **Generally lasts 20-30 minutes**
- **We highlight significant science results**
  - combustion
  - supernova simulations
  - climate
  - biochemistry
  - fusion
  - ...

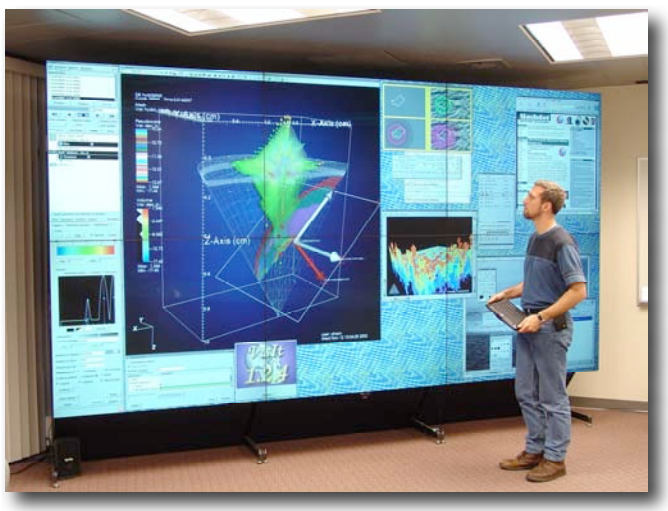


- **We would LOVE to include your work and be your mouthpiece**

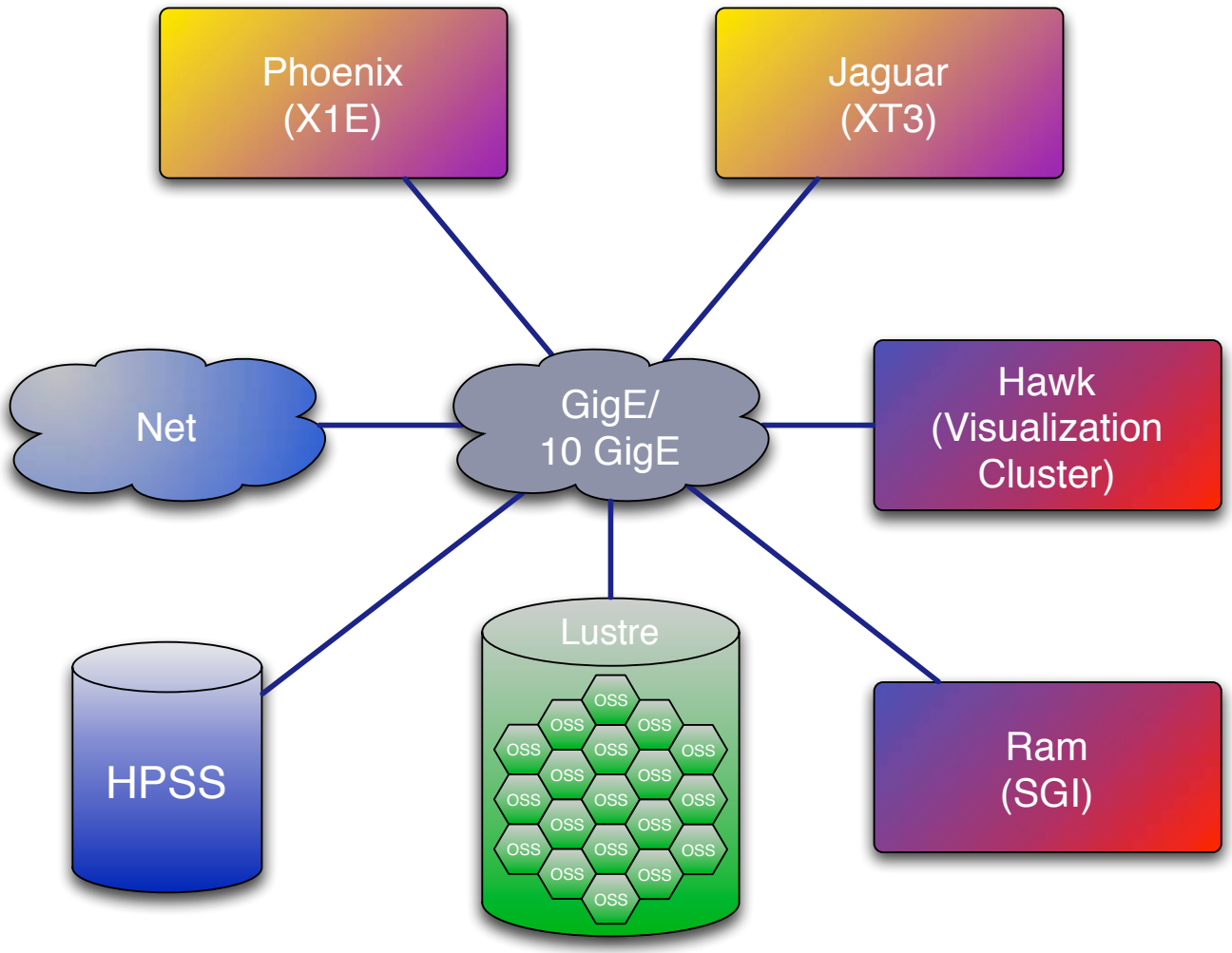
EVEREST

- Large format powerwall
- 30’ x 8’ in size
- 27 projectors
- 1280 x 1024 each: ~35 million pixels
- Able to run interactive visualization tools





Network Infrastructure



Deployed Major "Turnkey" tools

- **VisIt**
- **EnSight**
- **Paraview**
- **VMD**
- **AVS/Express**

- **What I mean by "turnkey" is:**
  - Rich set of features for visualization and analysis
  - No programming is required to access main features
  - Flexible in data input capabilities
  - The first three will scale to extremely large data

Getting access

- **Most tools accessible through "modules":**

```
% module avail
AVS-Express/7.0          ferret/5.81              netcdf/3.6.1
AVS-Express+PST/7.0      firefox/1.5              paraview/2.4-mpich
DMX-chromium             freeglut/2.4.0           povray/3.5.0c
DefApps                  glut/3.6                 povray/3.6
MiscApps                 glut/3.7                 povray/3.6.1
R/2.3.0                  gnuplot/4.0              scirun/3.0.0
blockbuster              hdf5/1.6.5               toolkit/1.0
cg/1.4.0-4               idl/6.2                  toolkit/1.1
chromium/1.2-32          idl/6.3(default)         valgrind/3.1.1
chromium/1.7-32          java/jre-1.5.0.06        valgrind/3.2.3
ensight/8.0              ncl/4.2.0.a33            visit
ensight/8.2              netcdf/3.6.0
```

- **Some documentation at**
- **Contact us – many more capabilities possible**

File systems on hawk

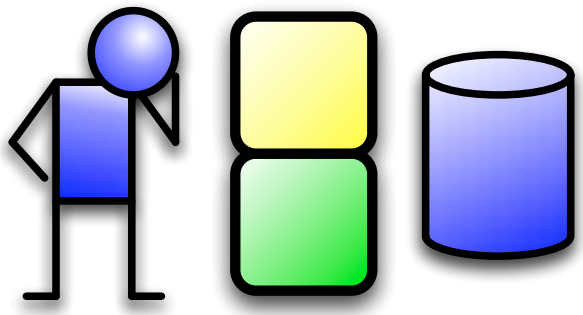
- **Your home directory: /spin/home/$USER**
  - Insignificant space (500 MB)
- **Scratch space: /scr2ta**
  - Small space: 1.5 TB
  - Pretty full, lot of contention
- **Dedicated NFS file system: /nfs/data/$USER**
  - Moderate space: 7.7 TB
  - Single server, so easy to saturate in parallel
- **(Coming soon) Lustre: /lustre/spider**
  - Good amount of space (80 TB)
  - Shared with Jaguar/Phoenix
  - Good parallel aggregate bandwidth

“Locality” Issues

- Data and customer are often not located together.
- Data is getting much larger.
- Data movement becoming increasingly painful.

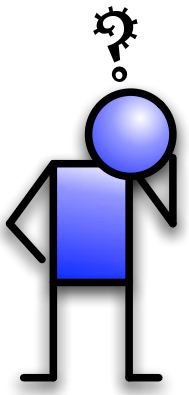
One solution: Decoupling of data processing from rendering/display

In the old days...

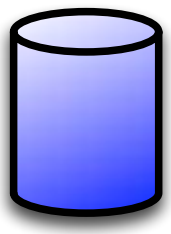


Now: Remote Customers…

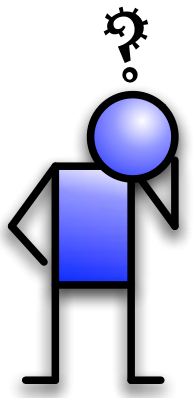
Your Site

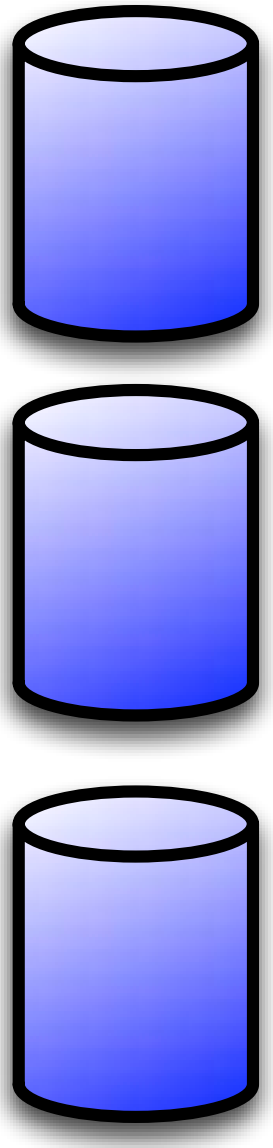


NCCS

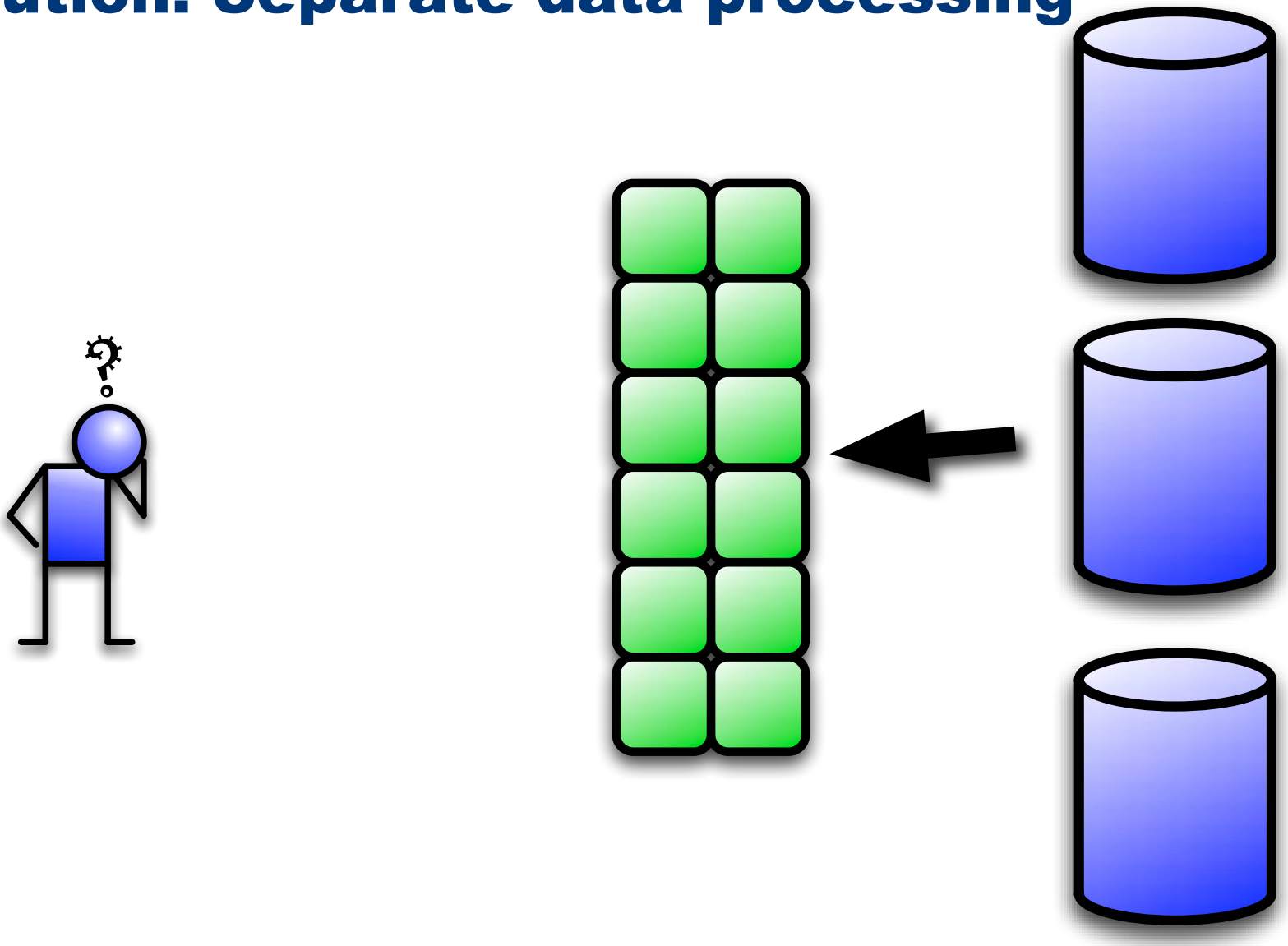


...with BIG data!

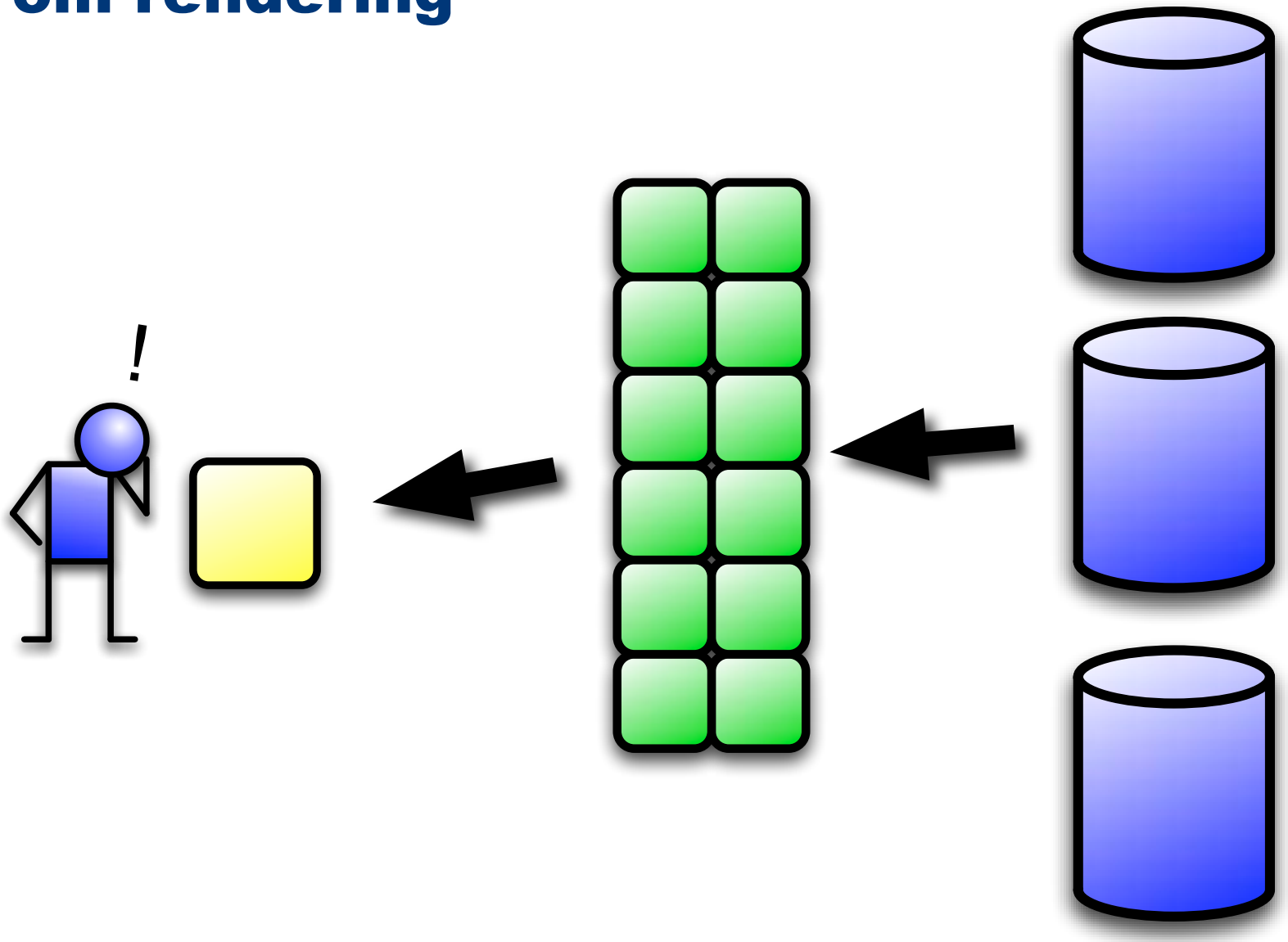




Solution: Separate data processing

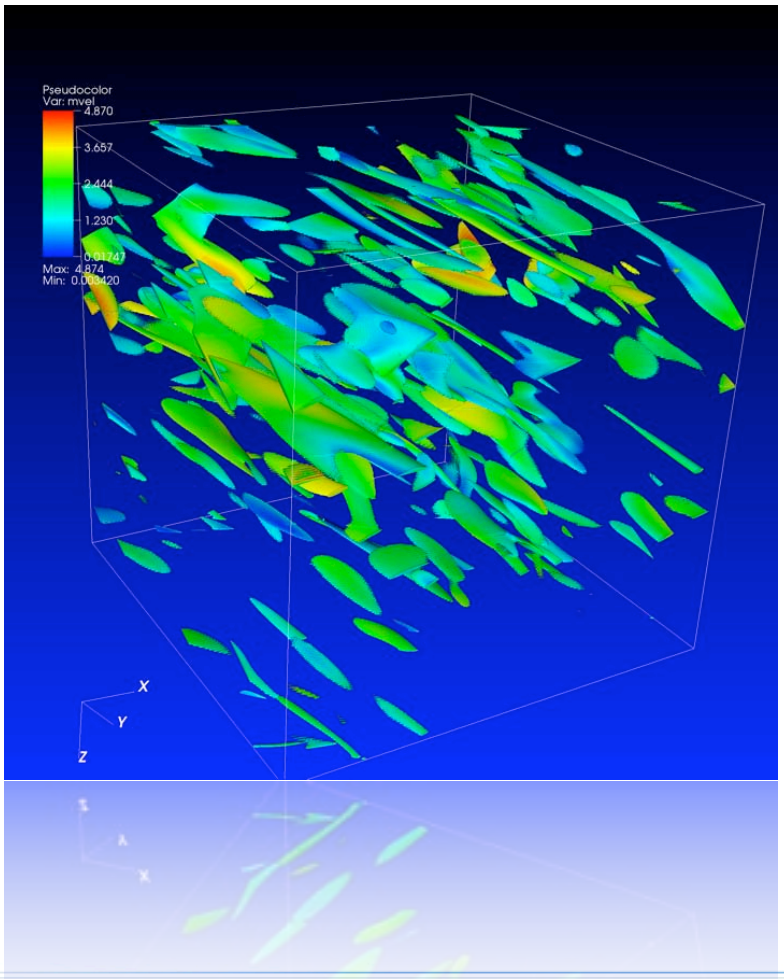


...from rendering

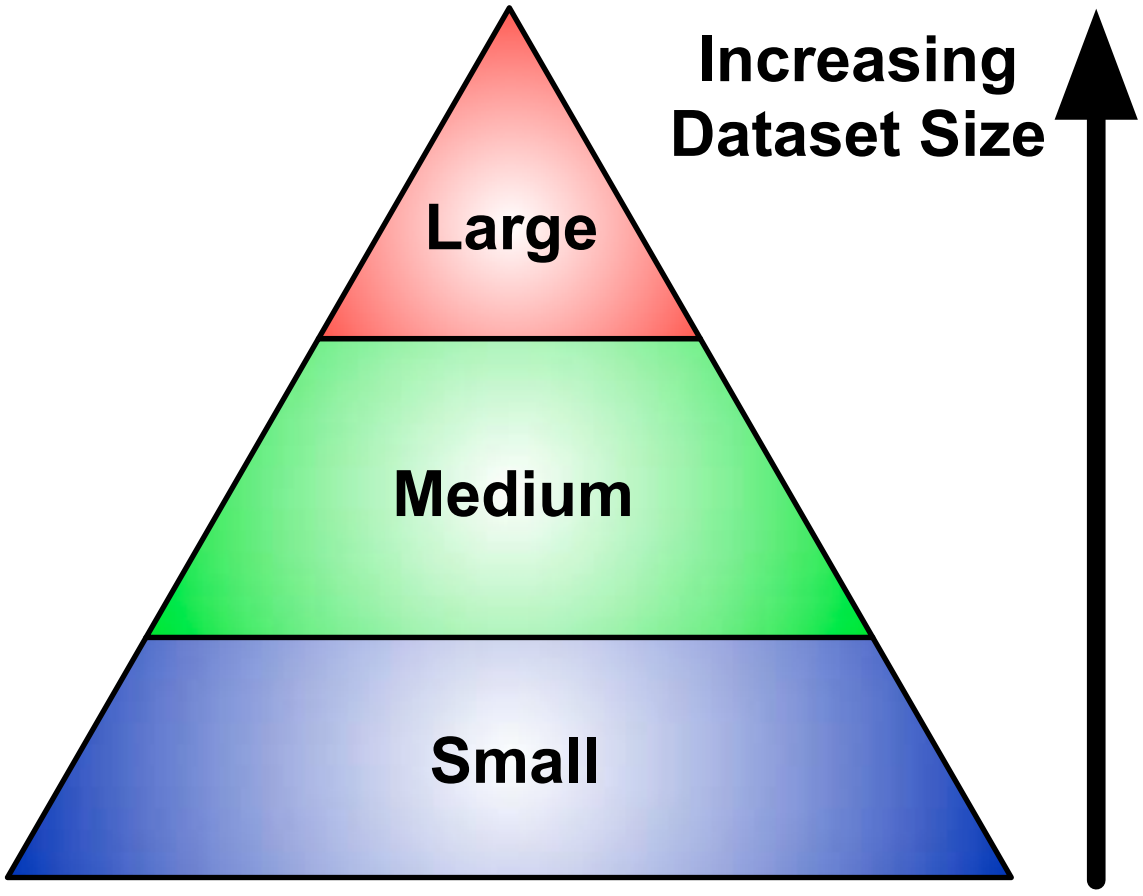


Remote Visualization

- **VisIt, Paraview, EnSight provide client/server access to large data visualization**
  - Geometry across the network
  - Imagery across the network
- **Issues/Concerns**
  - Latency
  - Socket initiation issues

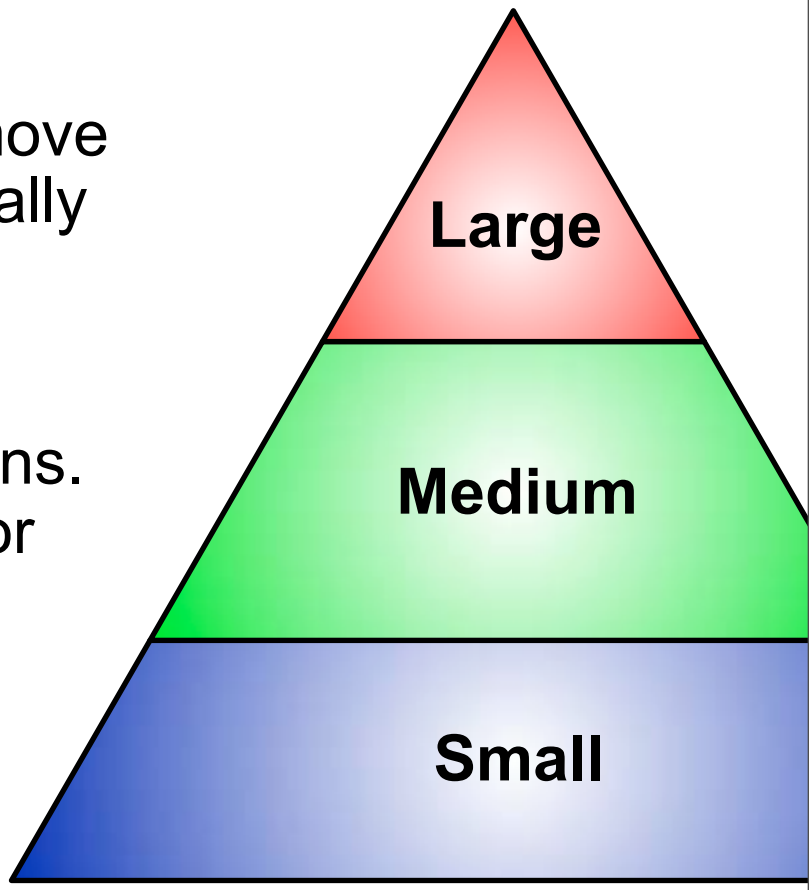


Dataset sizes change use cases



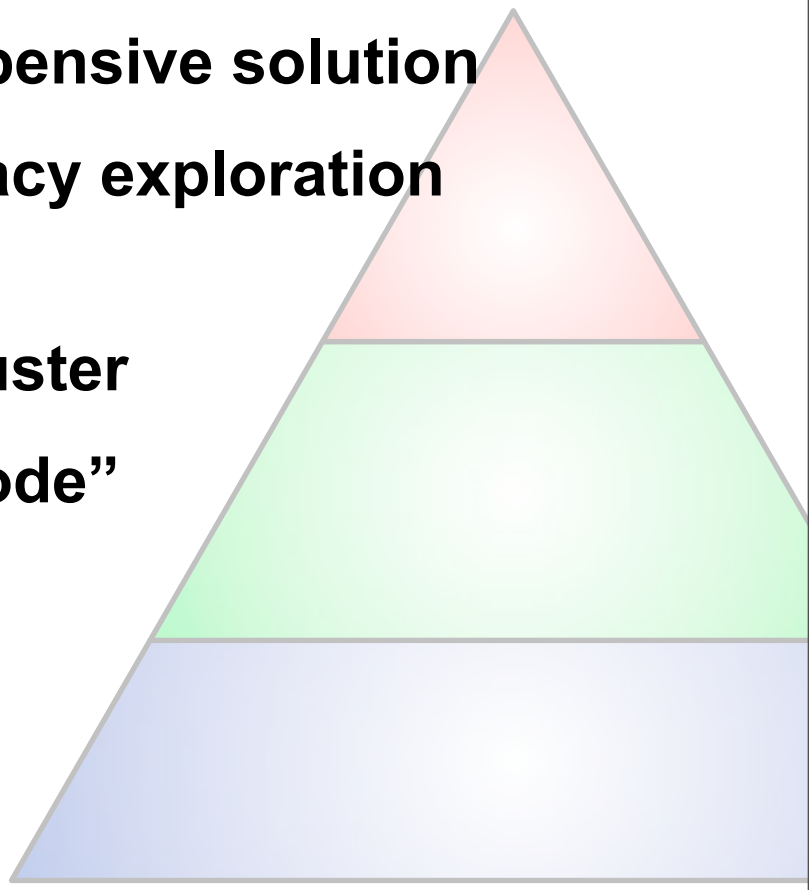
Data size categories

- **Small:**
  - Data is small enough to easily move anywhere. Analysis/vis is generally done on local workstations
- **Medium:**
  - Data won't fit on local workstations. Have to process on "fat" nodes or commodity clusters
  - Moderately painful to move.
- **Large:**
  - Data won't fit anywhere but the largest computational systems.
  - Functionally impossible to move.



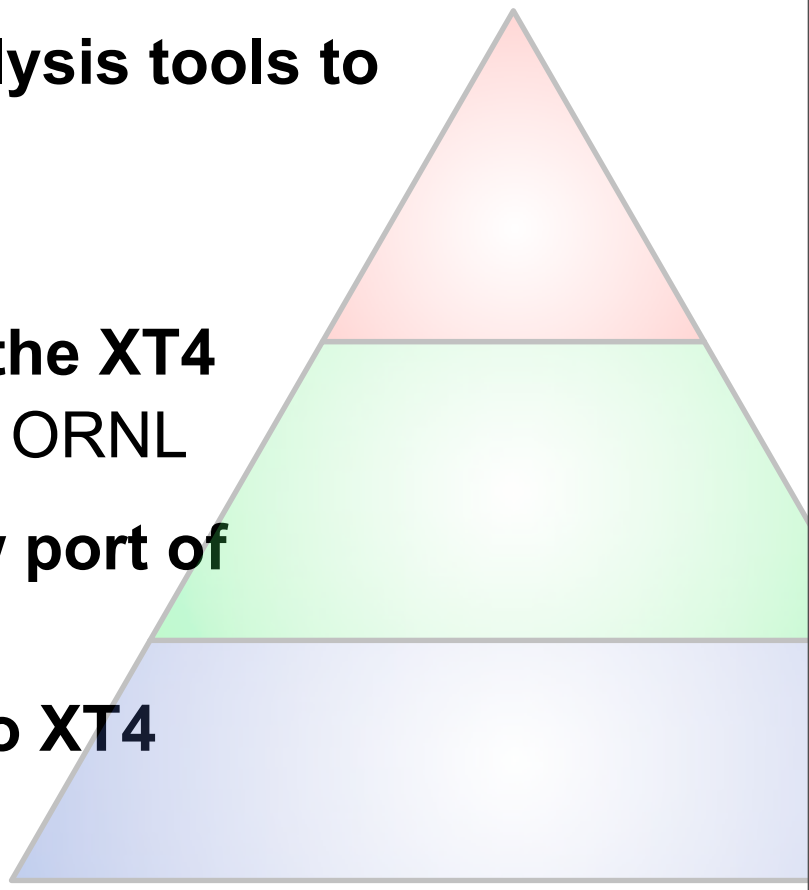
Medium-sized data

- **Commodity clusters are an inexpensive solution**
- **Large memory allows scalar/legacy exploration**
  - AVS, R, Tecplot, SCIRun, etc.
- **Currently served by the hawk cluster**
- **Now drafting RFP for new "fat node" cluster as a replacement.**



Largest data

- **Working to move largest vis/analysis tools to Cray XT4 architecture**
- **Cray has ported parallel VisIt to the XT4**
  - OS changes not yet deployed to ORNL
- **VisIt "lessons learned" will allow port of ParaView to XT4.**
- **Port EnSight Server-of-servers to XT4**



Visualization tutorial this afternoon

- **A "how-to" exploration of several visualization and analysis tools:**
  - NetCDF exploration
  - VisIt visualization system
  - R statistical analysis
- **Follow along at**
  - http://nccs.gov/news/workshops/fy07usersmeeting/visualization.html
- **May wish to download the climate dataset found there**
- **May wish to download VisIt:**
  - http://www.llnl.gov/visit/executables.html